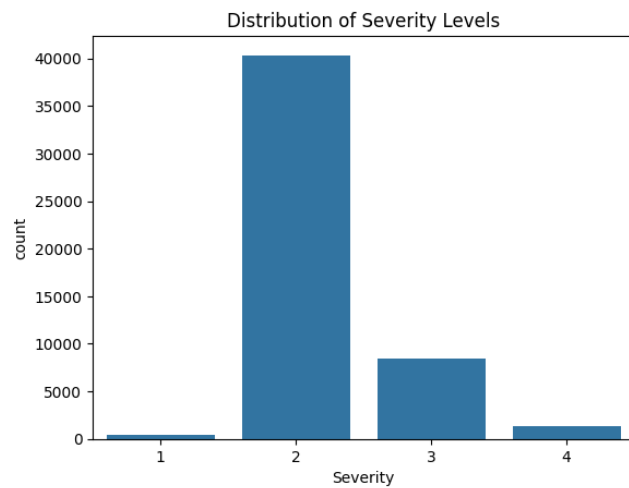
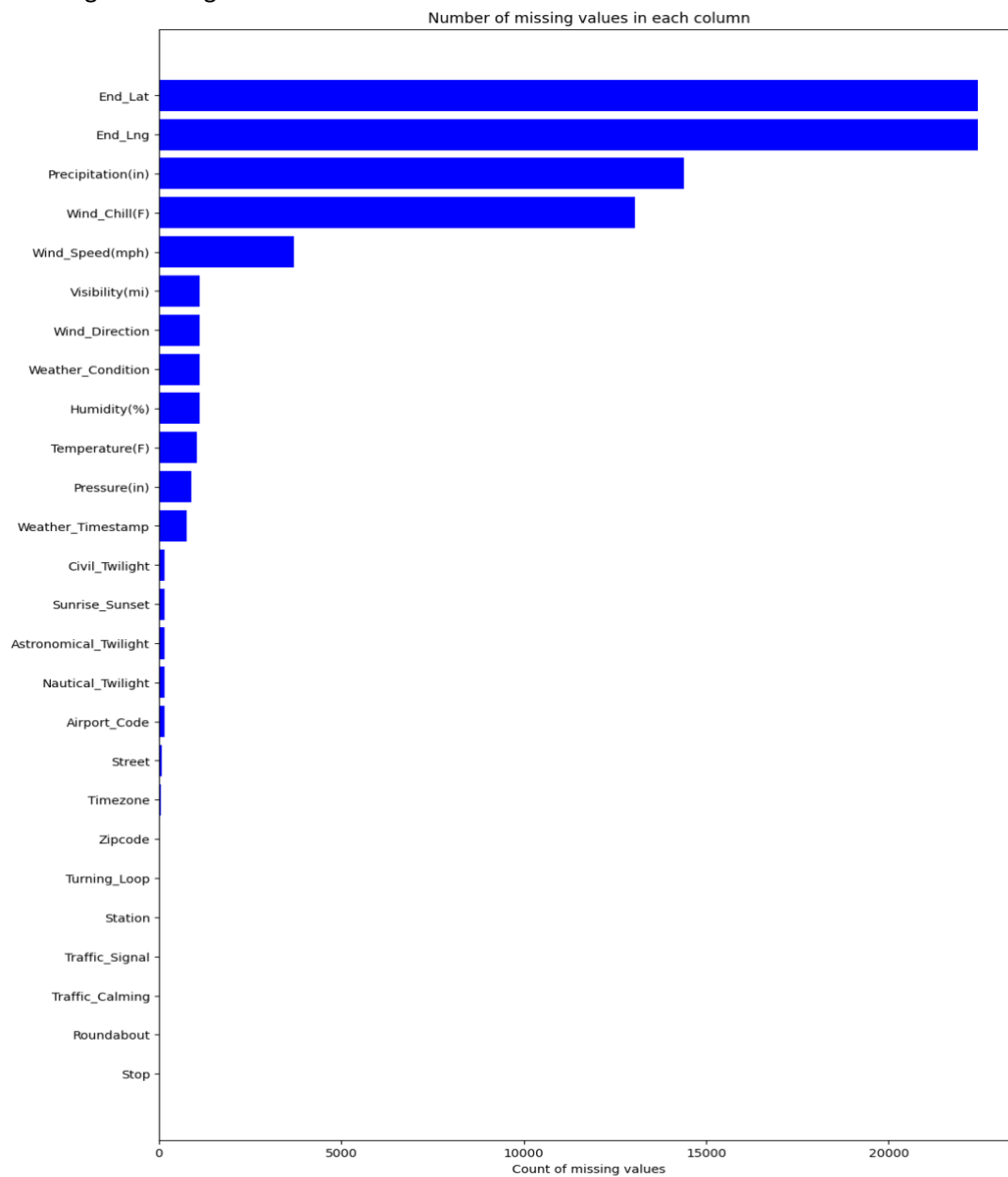


Results

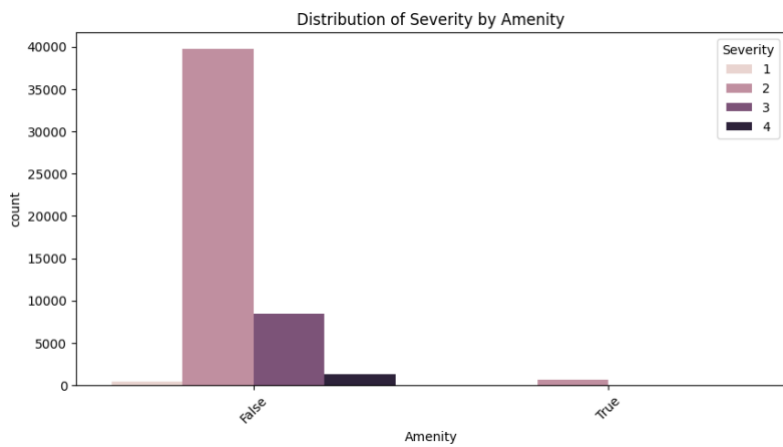
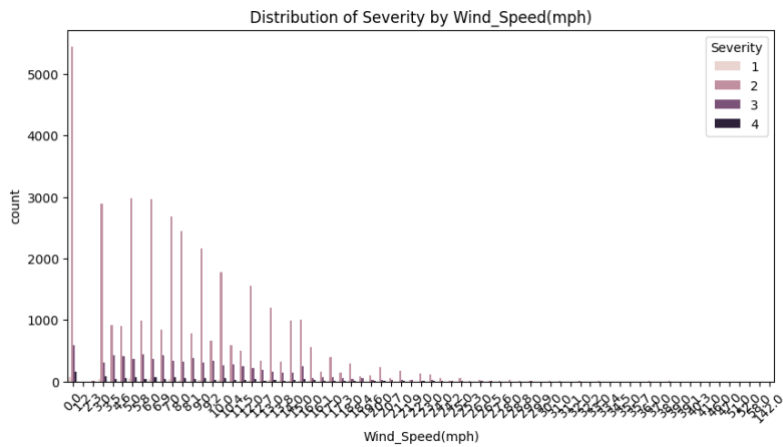
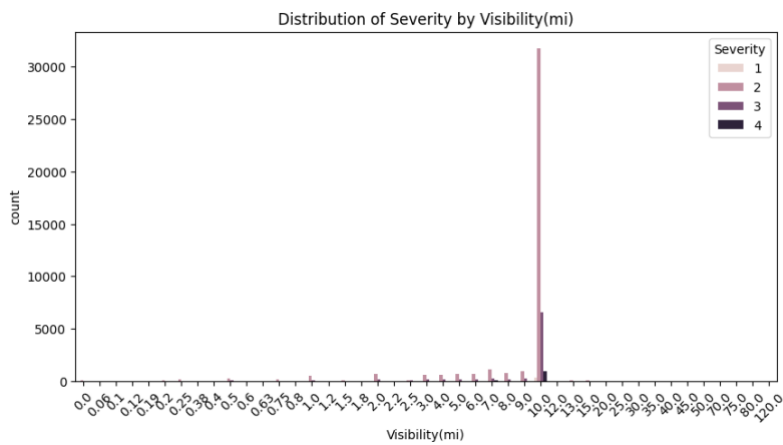
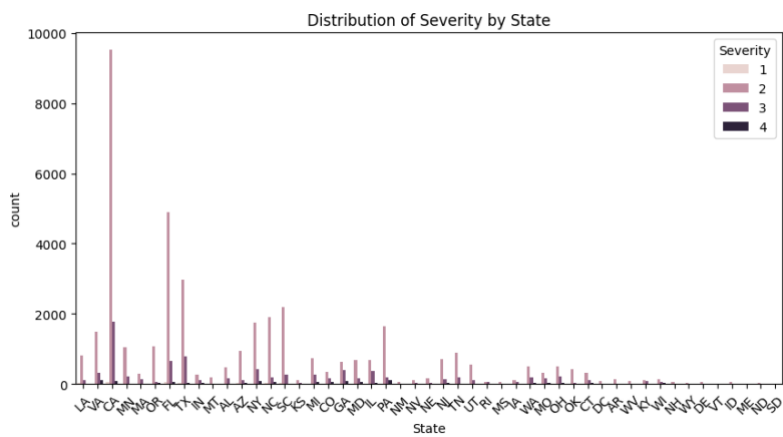
Distribution of Target Variable:

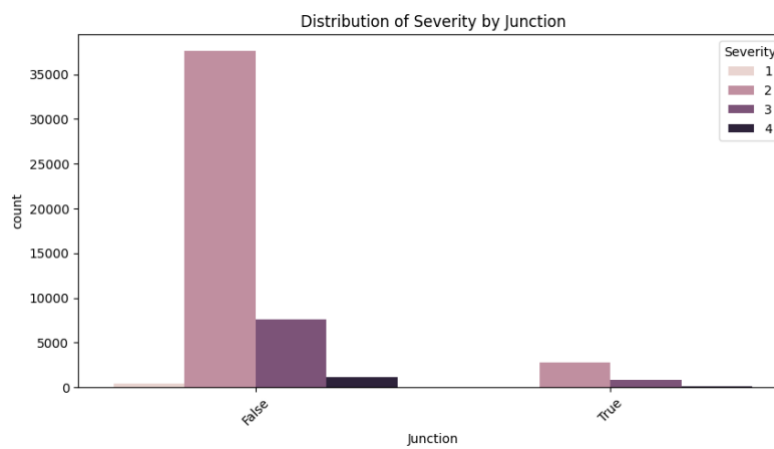
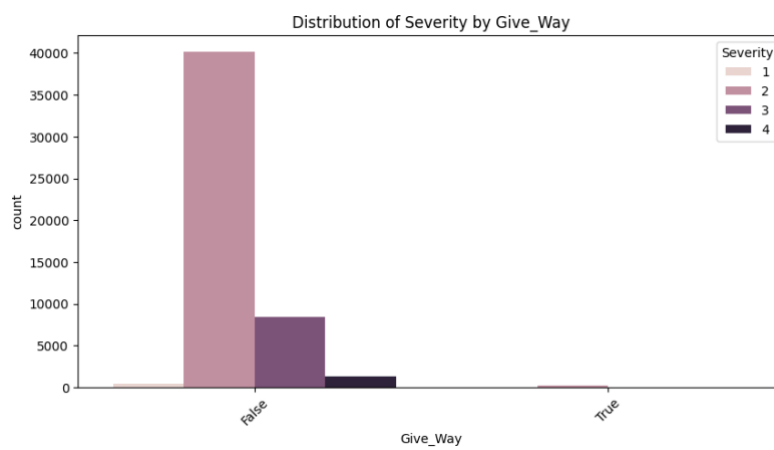
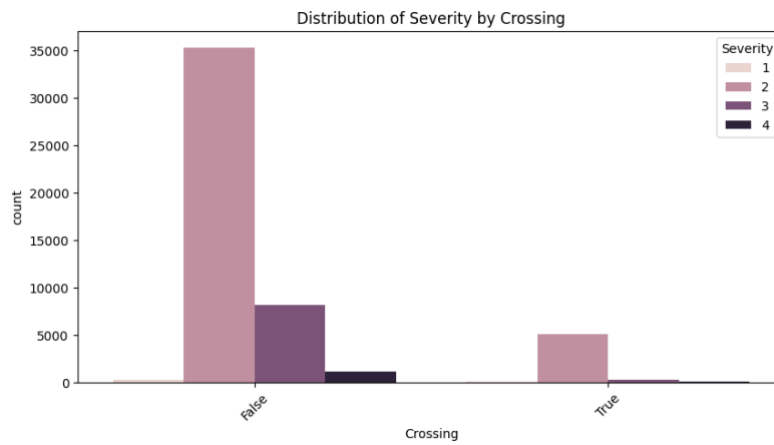
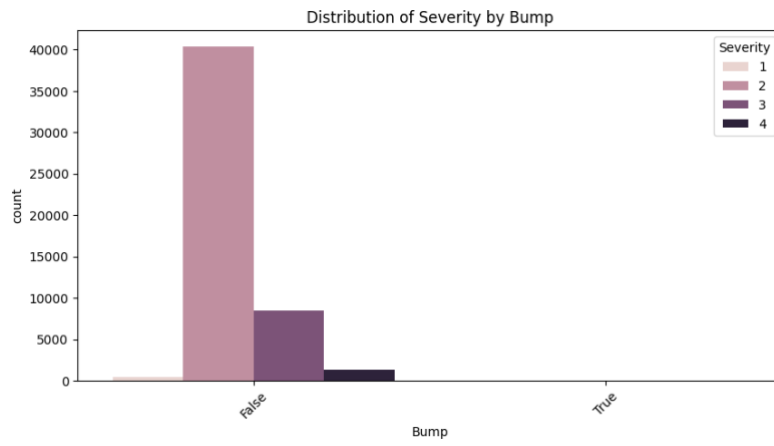


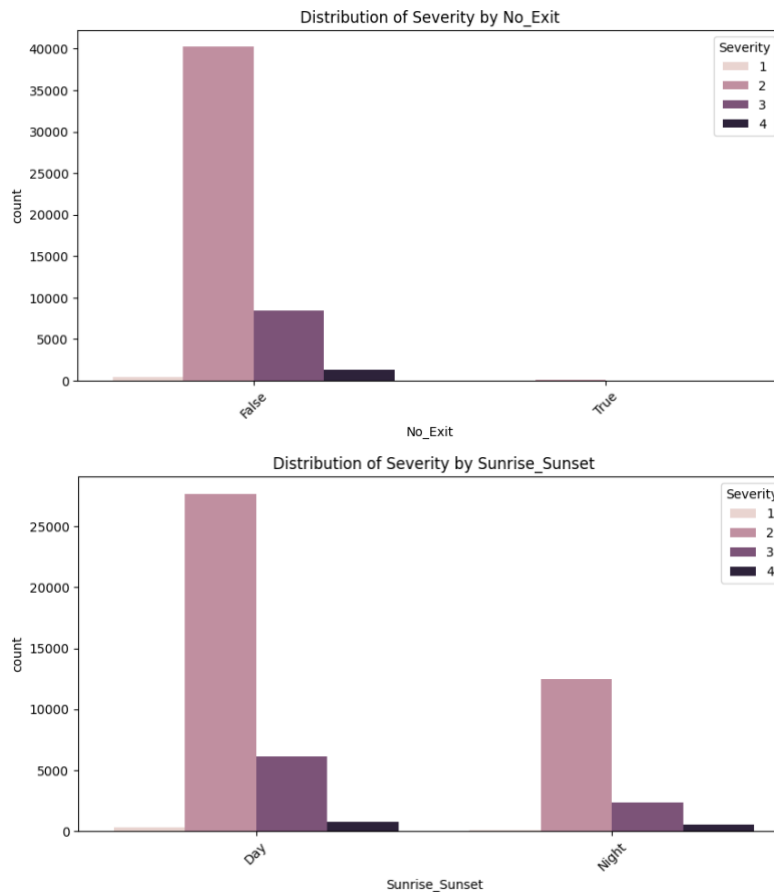
Cleaning of missing values from the dataset:



Research Question 1: Primary factors contributing to the “Severity” of traffic collisions in the United States?







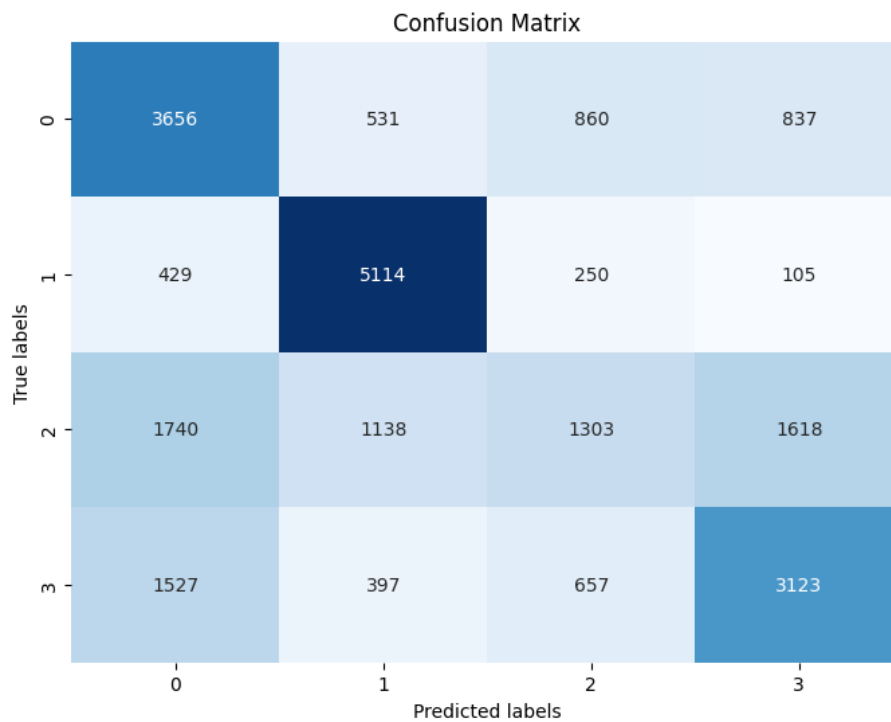
Research Question 2 and 3: Can machine learning models effectively classify collision severity based on historical data? How does the model's performance vary across different variants of ML algorithms?

After SMOTE oversampling and 80:20 train test split:

```
Shape of X_train: (93139, 68)
Shape of X_test: (23285, 68)
Shape of y_train: (93139,)
Shape of y_test: (23285,)
```

Model Evaluation after PCA:

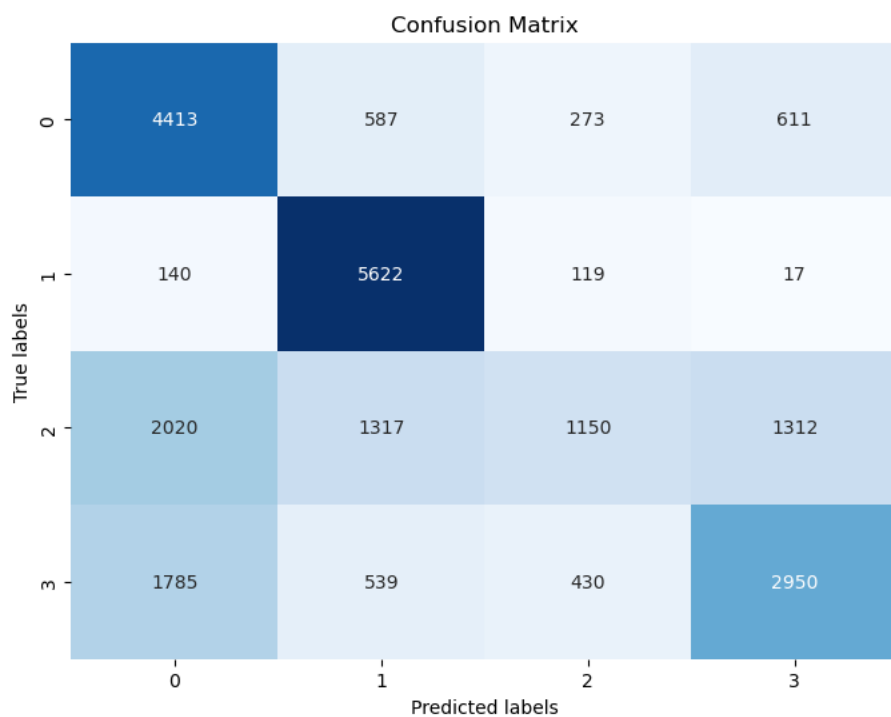
Logistic Regression:



Classification Report for Logistic Regression:

	precision	recall	f1-score	support
1	0.50	0.62	0.55	5884
2	0.71	0.87	0.78	5898
3	0.42	0.22	0.29	5799
4	0.55	0.55	0.55	5704
accuracy			0.57	23285
macro avg	0.55	0.57	0.54	23285
weighted avg	0.55	0.57	0.55	23285

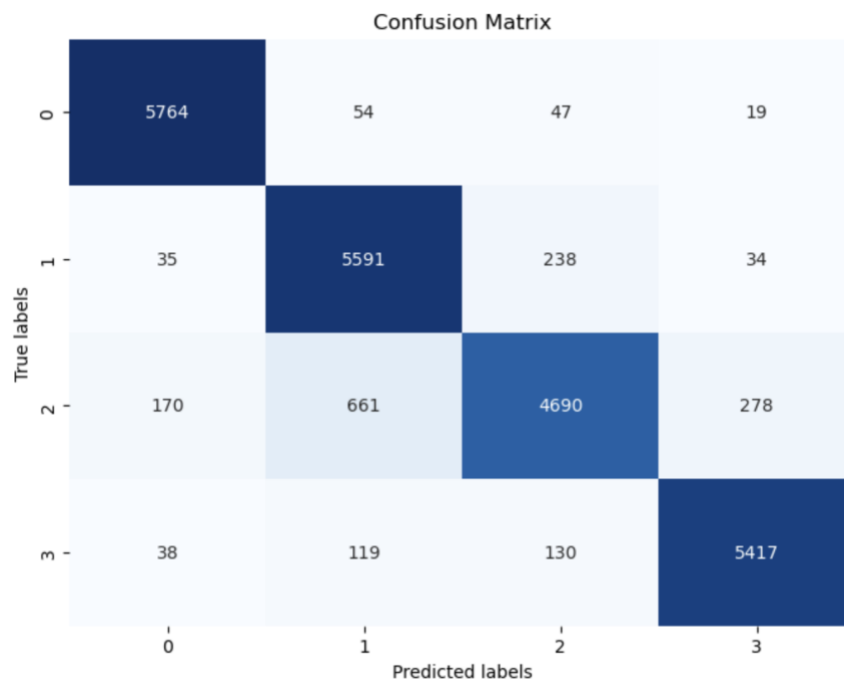
SVM:



Classification Report for SVM:

	precision	recall	f1-score	support
1	0.53	0.75	0.62	5884
2	0.70	0.95	0.81	5898
3	0.58	0.20	0.30	5799
4	0.60	0.52	0.56	5704
accuracy			0.61	23285
macro avg	0.60	0.60	0.57	23285
weighted avg	0.60	0.61	0.57	23285

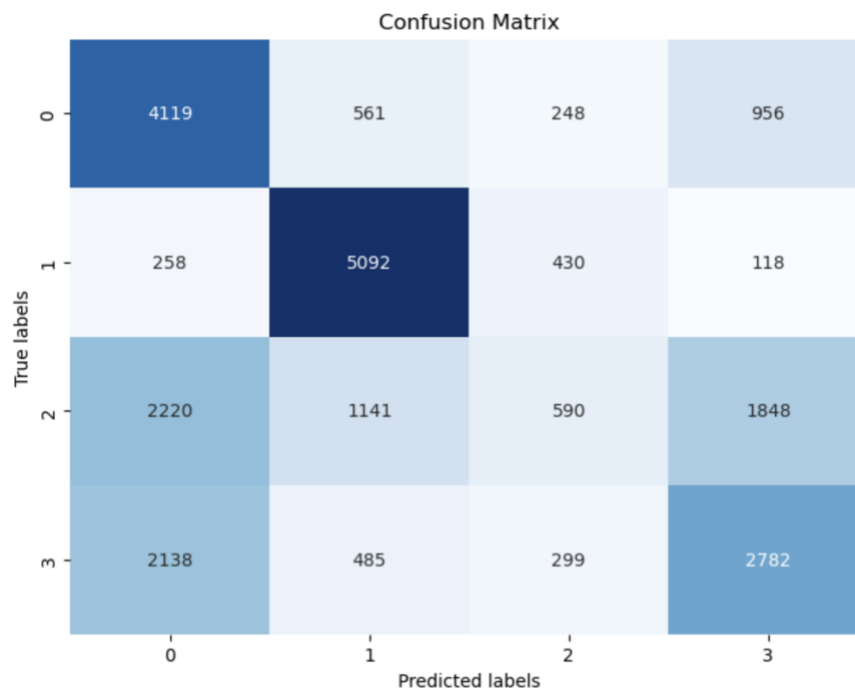
Random Forests:



Classification Report for Random Forest:

	precision	recall	f1-score	support
1	0.96	0.98	0.97	5884
2	0.87	0.95	0.91	5898
3	0.92	0.81	0.86	5799
4	0.94	0.95	0.95	5704
accuracy			0.92	23285
macro avg	0.92	0.92	0.92	23285
weighted avg	0.92	0.92	0.92	23285

Adaboost:



Classification Report for Adaboost:

	precision	recall	f1-score	support
1	0.47	0.70	0.56	5884
2	0.70	0.86	0.77	5898
3	0.38	0.10	0.16	5799
4	0.49	0.49	0.49	5704
accuracy			0.54	23285
macro avg	0.51	0.54	0.50	23285
weighted avg	0.51	0.54	0.50	23285